FRONTIERS IN GENETICS

SPECIALTY GRAND CHALLENGE ARTICLE

published: 01 March 2011

doi: 10.3389/fgene.2011.00009

Grand challenge in behavioral and psychiatric genetics: quantitative challenges to keeping up with molecular advances

Valerie S. Knopik*[1,2]

[1] *Division of Behavioral Genetics, Department of Psychiatry, Rhode Island Hospital, Providence, RI, USA*

[2] *Department of Psychiatry and Human Behavior, Warren Alpert School of Medicine, Brown University, Providence, RI, USA*

**Correspondence: valerie_knopik@brown.edu*

The historical split between the worlds of quantitative and molecular genetics happened over a century ago and was born of divided emphasis on naturally occurring variation in complex traits and species-typical phenomena (i.e., assuming all members of a species are genetically the same except for a few rogue mutations that disrupt normal processes), respectively (Plomin et al., 2003). This split allowed independent progress to be made in these two different disciplines. Quantitative genetics, through family based designs and animal models considering naturally occurring genetic variation in mice, for example, informed much of what we know about heritability today. Molecular genetics, in contrast, asked whether manipulation of the genetic code, such as "knocking out" a gene, could have an average effect on organisms via altered regulation, under- or over-expression of a gene. Efforts to reconcile the drifting apart of these two approaches began in the 1980s with the advent of DNA markers that, because they were polymorphisms in the DNA itself rather than in a gene product (i.e., red cell blood proteins) held promise for identifying quantitative trait loci (QTLs) responsible for the inheritance of complex traits, such as those in behavioral and psychiatric genetics. However, progress has been slower than expected. In fact, more recent molecular advances point to how quantitative and bioinformatic approaches are lagging behind. This recent advent of new advanced molecular technologies has provided an efficiency in genotyping that allows these two complementary fields of study to synergize to a greater degree.

The pace at which information about genetic and epigenetic modification is being produced has created a strain on our available analytic tools when we try to consider the impact or significance of this genetic information. A myriad of molecular techniques exist to characterize rare and common variants, copy number variants (CNVs), and even phase (i.e., the combination of alleles specific to each parental chromosome). This poses a significant challenge to the field as far as integrating the effects of these individual differences to better understand the biological basis of behavior. For example, Lipsky et al. (2009) reported on a number of variants in SLC6A4 (the serotonin transporter gene) thought to impact expression, yet studies are still being published with only biallelic characterization of 5HTTLPR (a polymorphic region within SLC6A4). Initial evidence suggests that DNA methylation of SLC6A4 (Beach et al., 2011) may also be important. The laboratories publishing biallelic data are frequently not lacking in the technology to assess these additional genetic and epigenetic variants, but combining this information analytically without a tremendous cost of power may instead be a limiting factor. This points to a challenge toward the efficient integration of the increasing amount of biological information available.

An additional challenge is to leverage biological knowledge in the search for genes related to etiology. The agnostic approach used by genome-wide association studies (GWAS) has led to important breakthroughs, but the enormous multiple comparison penalty incurred by this, and similar, methods have proven to be a significant hurdle. A complementary approach that may be useful is to prioritize genes in biological pathways known to underlie the phenotype of interest. Such an approach can capitalize on the economy of genotyping afforded by new technology without necessarily paying the statistical penalty for each and every potential comparison. Curated databases of phenotype-specific collections of genes can provide a narrowed focus for investigating biological systems. Examples of these databases are described in the study of cervical cancer and the prostate (Li et al., 2003; Agarwal et al., 2011). These databases provide a gene set that can be examined in individuals with genome-wide genotyping to obtain the financial advantages of this technology while simultaneously increasing the likelihood of finding a signal by decreasing the number of comparisons. It is important to remember that such an approach would be in addition to agnostic strategies, as this approach assumes a comprehensive biological knowledge that is unlikely to be the case in behavioral and psychiatric phenotypes. The evolving definition of systems biology also includes the merging of bioengineering principles with measured genotypes in order to begin to consider new methods for aggregating risk. For example, perturbations in neurotransmitter systems may only become evident when multiple loci are considered within the system concurrently. These variants may have additive or even multiplicative effects when considered systematically and may account for substantially more variance than present analytic methods would suggest.

Adding difficulty to this already Herculean task is the fact that psychiatric and behavioral phenotypes are often the result of gene products expressed in brain. That is, these phenotypes cannot be observed directly and are measured by psychometric or diagnostic instruments, which are imprecise. Since this tissue is very rarely available in living human individuals, the advantages afforded to the study of other non-brain related medical conditions (e.g., characterization of gene expression and epigenetic modification in target tissues) are not available to our field. Preclinical models will provide much needed guidance to address this issue of limited access to target tissues, but building readily translatable models provides its own challenges. This is

not to say that there is no utility in studying these processes in human peripheral tissues. Discovery of a peripheral biomarker would have clinical utility inasmuch as those peripheral tissues are available to clinicians to aid in diagnosis and treatment planning. However, the inability to access the tissue of interest presents an additional challenge to the field relative to other areas of genetic inquiry (e.g., oncogenetics). It seems likely that the differences in gene expression and epigenetic modification would be specific not only to a brain region, but to a specific nuclei within that region. If this is true, advances in psychiatric and behavioral genetics will almost certainly lag behind areas of inquiry into accessible tissues. Neuroimaging techniques that allow characterization of gene product levels, for example PET, may be of utility in this case as the protein product is presumably the ultimate consequence of gene expression differences and epigenetic modification.

Outcomes such as neuroimaging variables indicate the very important need of careful phenotypic characterization. Using attention-deficit hyperactivity disorder (ADHD) as an example, how can we as a field, reconcile the fact that the variance explained by successfully located genetic variants for ADHD account for only 1% of the variance (Franke et al., 2009), while heritability estimates from twin and family studies suggest ~80% (Knopik et al., 2005)? This concept of "missing heritability," i.e., the presumed inconsistency between heritability estimates and the ability to explain this genetic variance with genetic markers (Maher, 2008) is non-trivial. Indeed, a recent publication would suggest that measured genetic variants do not improve upon the predictive power of family history alone when it comes to cardiovascular disease (Paynter et al., 2010). Many explanations have been put forward including, but not limited to, limitations of genetic models, and statistical techniques commonly used in association studies (de los Campos et al., 2010), gene–environment interaction, epistasis, epigenetics, penetrance, rare variants, incomplete coverage of the gene, and genetic heterogeneity (see van der Sluis et al., 2010). Advances in molecular and statistical techniques will assist on one end of this problem; however, "at least as important to the detection of genetic variants for complex traits is the way complex traits are measured and the phenotypic information is modeled" (van der Sluis et al., 2010). In their recent report, van der Sluis et al. (2010) elegantly show how phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem and conclude that careful phenotypic data modeling can improve the genetic signal, leading to identification of genetic variants that may otherwise go undetected, and increase statistical power by 20–99%.

This phenotypic refinement, however, invites the question: How can we balance the need for careful phenotypic characterization with the need for increasingly large sample sizes which will protect against false-positive results from genome scans or deep-sequencing? Can we adopt a synergistic approach that does not necessarily imply that the only way to do meaningful genetically informed research in behavioral and psychiatric outcomes, is to be part of a large consortium? This option, while certainly important, is not viable or appropriate for all research questions. When intensive phenotyping results in a diminished sample size (e.g., neuroimaging) where the number of genetic markers far exceeds the number of participants, can we capitalize on the rich history of agricultural, and livestock genetics by adopting such approaches as whole-genome marker-enabled prediction (WGP) as discussed in de los Campos et al. (2010)? Do we invest our efforts in WGP and alternative aggregate genetic risk approaches (e.g., Cornelis et al., 2009; DeJager et al., 2009; Purcell and The International Schizophrenia Consortium, 2009; Belsky and Beaver, 2010) and extending these beyond using one SNP per gene to include haplotypes, account for linkage disequilibrium, and differences in inheritances patterns on a locus-by-locus basis? How can we capitalize on recent advances that generate completely phased genome sequence information (Yang et al., 2011) to advance the study of heritable disorders? Can we also consider the use of a combination of approaches or perhaps an aggregate approach, including information from human (family, case–control, independent individuals) and animal models (Patel et al., 2010), gene networks (Ruano et al., 2010), and systems biology to gain traction? A daunting task is to design research powerful enough to integrate the advances in molecular genetics to find meaningful and robust results as we search for genes involved in etiology of behavioral and psychiatric outcomes and move toward the possibility of personalized medicine.

A final challenge to our field is to recall that behavioral and psychiatric genetic research not only tell us about genetics but it also tells us about environment. In fact, they provide the best possible evidence for the importance of the environment. That is, if the heritability of a certain trait is 50%, the remaining 50% is due to the environment. Genes do not equal destiny. The environment impacts the development of behavioral and psychiatric disorders and the environment influences genetics as well (Li, 2010). There is a clear need to carefully measure the environment and to be mindful that many environmental measures show genetic influence (Kendler and Baker, 2007). Thus, genetically sensitive designs that incorporate and investigate interactions and correlations between genes and environments are key.

In sum, behavioral and psychiatric genetics is a fairly young field that is experiencing an explosion of growth. Advances in molecular analytic processing present heretofore unseen possibilities for understanding the genetic contribution to behavioral and psychiatric outcomes. It is our job to begin to leverage biological knowledge in the search for genes related to etiology, and to develop new and creative research designs and techniques to integrate the vast amount of biological information into models that also include highly refined phenotypes and careful measurement of the environment. This will, necessarily, have to be a team science approach – one that includes contributions from clinicians, bioengineers, bioinformaticians, statistical geneticists, molecular geneticists, psychometricians, computational biologists, anthropologists, evolutionary biologists, among others. It is "early days," yet exciting days, and I look forward to working alongside all of you as we embrace these challenges together.

REFERENCES

Agarwal, S. M., Raghav, D., Singh, H., and Raghava, G. P. (2011). CCDB: a curated database of genes involved in cervix cancer. *Nucleic Acids Res.* 39, D975–D979.

Beach, S. R., Brody, G. H., Todorov, A. A., Gunter, T. D., and Philibert, R. A. (2011). Methylation at 5HTT mediates the impact of child sex abuse on women's

antisocial behavior: an examination of the Iowa adoptee sample. *Psychosom. Med.* 73, 83–87.

Belsky, J., and Beaver, K. M. (2010). Cumulative-genetic plasticity, parenting and adolescent self-regulation. *J. Child Psychol. Psychiatr.* doi: 10.1111/j.1469-7610.2010.02327.x. [Epub ahead of print].

Cornelis, M. C., Qi, L., Zhang, C., Kraft, P., Manson, J., Cai, T., Hunter, D. J., and Hu, F. (2009). Joint effect of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. *Ann. Intern. Med.* 150, 541–550.

de los Campos, G., Gianola, D., and Allison, D. B. (2010). Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nat. Rev. Genet.* 11, 880–886.

DeJager, P., Chibnik, L. B., Cui, J., Reischl, J., Lehr, S., Simon, K. C., Aubin, C., Bauer, D., Heubach, J. F., Sandbrink, R., Tyblova, M., Lelkova, P.; The Steering Committees of BENEFIT, BEYOND, LTF, and CCR1 Studies. (2009). Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Neurology* 8, 1111–1119.

Franke, B., Neale, B. M., and Faraone, S. V. (2009). Genome-wide association studies in ADHD. *Hum. Genet.* 126, 13–50.

Kendler, K. S., and Baker, J. H. (2007). Genetic influences on measures of the environment: a systematic review. *Psychol. Med.* 37, 615–626.

Knopik, V. S., Sparrow, E. P., Madden, P. A., Bucholz, K. K., Hudziak, J. J., Reich, W., Slutske, W. S., Grant, J. D., McLaughlin, T. L., Todorov, A., Todd, R. D., and Heath, A. C. (2005). Contributions of parental alcoholism, prenatal substance exposure, and genetic transmission to child ADHD risk: a female twin study. *Psychol. Med.* 35, 625–635.

Li, L. C., Zhao, H., Shiina, H., Kane, C. J., and Dahiya, R. (2003). PGDB: a curated and integrated database of genes related to the prostate. *Nucleic Acids Res.* 31, 291–293.

Li, M. D. (2010). Grand challenges and opportunities for molecular psychiatry research: a perspective. *Front. Psychiatr.* 1:2. doi: 10.3389/fpsyt.2010.00002

Lipsky, R. H., Hu, X. Z., and Goldman, D. (2009). Additional functional variation at the SLC6A4 gene. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 150B, 153.

Maher, B. (2008). Personal genomes: the case of the missing heritability. *Nature* 456, 18–21.

Patel, S. D., Le-Niculescu, H., Koller, D. L., Green, S. D., Lahiri, D. K., McMahon, F. J., Nurnberger, J. I. Jr., and Niculescu, A. B. III. (2010). Coming to grips with complex disorders: genetic risk prediction in bipolar disorder using panels of genes identified through convergent functional genomics. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 153B, 850–877.

Paynter, N. P., Chasman, D. I., Paré, G., Buring, J. E., Cook, N. R., Miletich, J. P., and Ridker, P. M. (2010). Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 303, 631–637.

Plomin, R., DeFries, J. C., Craig, I. W., and McGuffin, P. (2003). *Behavioral Genetics in the Postgenomic Era*. Washington, DC: American Psychological Association.

Purcell, S., and The International Schizophrenia Consortium. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748–752.

Ruano, D., Abecasis, G. R., Glaser, B., Lips, E. S., Cornelisse, L. N., de Jong, A. P., Evans, D. M., Davey Smith, G., Timpson, N. J., Smit, A. B., Heutink, P., Verhage, M., and Posthuma, D. (2010). Functional gene group analysis reveals a role of synaptic heterotrimeric G proteins in cognitive ability. *Am. J. Hum. Genet.* 86, 113–125.

van der Sluis, S., Verhage, M., Posthuma, D., and Dolan, C. V. (2010). Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. *PLoS ONE* 5, e13929. doi: 10.1371/journal.pone.0013929

Yang, H., Chen, X., and Wong, W. H. (2011). Completely phased genome sequencing through chromosome sorting. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12–17.

Received: 17 February 2011; accepted: 17 February 2011; published online: 01 March 2011.

Citation: Knopik VS (2011) Grand challenge in behavioral and psychiatric genetics: quantitative challenges to keeping up with molecular advances. Front. Gene. ***2****:9. doi: 10.3389/fgene.2011.00009*

This article was submitted to Frontiers in Behavioral and Psychiatric Genetics, a specialty of Frontiers in Genetics.